



(12) **EUROPEAN PATENT APPLICATION**
published in accordance with Art. 153(4) EPC

(43) Date of publication:
10.11.2021 Bulletin 2021/45

(51) Int Cl.:
C12Q 1/6806 (2018.01)

(21) Application number: **19907784.3**

(86) International application number:
PCT/CN2019/123891

(22) Date of filing: **09.12.2019**

(87) International publication number:
WO 2020/140693 (09.07.2020 Gazette 2020/28)

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR
Designated Extension States:
BA ME
Designated Validation States:
KH MA MD TN

(72) Inventors:
• **GUO, Zhiwei**
Shanghai 201210 (CN)
• **LI, Yinghui**
Shanghai 201210 (CN)
• **CHEN, Qian**
Shanghai 201210 (CN)
• **HU, Rongjun**
Shanghai 201210 (CN)

(30) Priority: **02.01.2019 CN 201910002408**
20.09.2019 CN 201910897689

(74) Representative: **J A Kemp LLP**
80 Turnmill Street
London EC1M 5QU (GB)

(71) Applicant: **Shanghai Zenisight Ltd.**
Shanghai 201210 (CN)

(54) **GENE TARGET REGION ENRICHMENT METHOD AND KIT**

(57) Provided is a gene target region enrichment method and a kit. The method comprises (1) amplifying fragmented DNA comprising a target region by means of a specific probe so as to obtain a captured-extension product, wherein the specific probe comprises a sequence complementary to the target region of the frag-

mented DNA, and the 3' terminal nucleotide of the specific probe is modified to prevent a ligation reaction at the 3' terminal of the specific probe; and (2) linking the 3' terminal of the captured-extension product obtained in step (1) to linker DNA to obtain a ligation product.

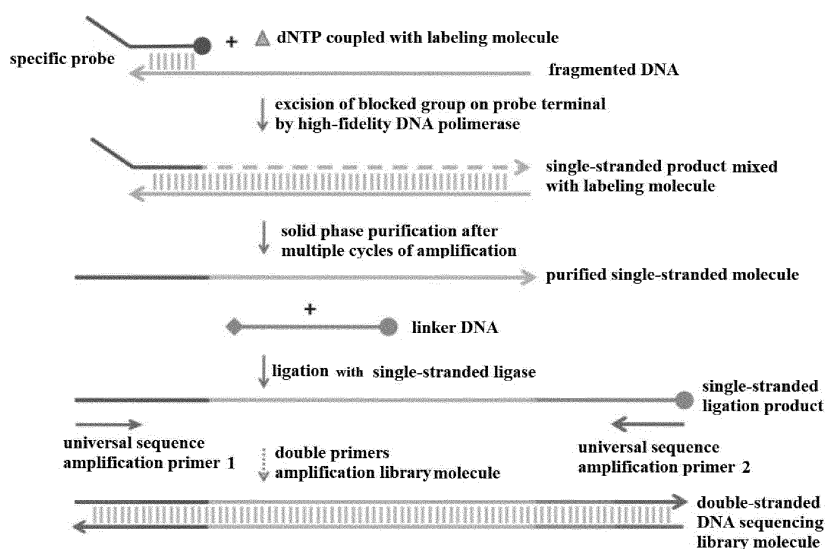


FIG.1

Description**TECHNICAL FIELD**

5 **[0001]** The present invention relates to the field of biotechnology, in particular, to a gene target region enrichment method and kit.

BACKGROUND OF THE INVENTION

10 **[0002]** Gene sequencing technology has gone through nearly half a century since it appears in the 1970s. The emergence of PCR technology in 1985 has promoted the development of the entire field of molecular biology. Next-generation sequencing technology (NGS) has the advantages of accuracy, sensitivity, and high throughput. As the cost of sequencing continues to decrease, its application range continues to expand, but its application is also subject to diverse requirement and time-consuming laborious library construction. In the process of building clinical samples such as plasma samples,
15 the existing forms of DNA are usually short fragments, damaged, single-stranded or partially double-stranded. For these existing forms, especially DNA with fragments less than 200bp, the present PCR technology cannot achieve good capture and enrichment.

[0003] For the enrichment of tiny DNA fragments, the existing technology still mainly uses the traditional PCR library construction method or the method of first adding a linker and then amplifying, such as the hybrid capture method.
20 However, for the former, due to the need for double-ended primers, the length of fragments suitable for amplification is greatly limited, and the preference in amplification results in high heterogeneity of the product, and errors accumulated by exponential amplification leads to subsequent sequencing result is inaccurate. While for the latter, although the requirement for the length of the enriched fragment is not as strict as PCR, the ligation reaction needs to be performed firstly, and the efficiency of ligation is usually only 20% to 50%, resulting in low capture efficiency, in addition, there is
25 also the problem of being easy to lose rare molecules due to difficulty in ligation. In order to solve the problem of NGS library construction, recently developed technologies include such as molecular inversion probes, multiple PCR and the like. Compared with hybrid capture technology, the molecular inversion probes have better specificity, but design of its pocket-shaped probe is complicated, and it is not suitable for the enrichment of tiny DNA fragments. Multiple PCR technology is suitable for large-scale samples and is the most widely used. But neither of them is suitable for enrichment
30 of small DNA fragments at low starting concentration, since either the requirements to primer design are extremely high and the homogeneity of the amplicon is poor, or the homogeneity of the amplified product is good but the requirement to the concentration of the starting sample is very high. These prior arts usually require double-ended primers to construct libraries, so, in order to remove linker dimer contamination, a purification step must be introduced, which results in the loss of information in small fragments of double-stranded DNA, damaged double-stranded DNA, and single-stranded
35 DNA molecules. However, in some genomic regions where transcription is active, it is precisely these forms of DNA. To sum up, the prior arts can't satisfy the requirement for enrichment of fragmented, especially small DNA of 200bp or less. Therefore, the main technical purpose of the present invention is to provide a method for efficient enrichment of the target regions in fragmented DNA, breaking the bottleneck of the low ligation efficiency limiting the enrichment efficiency, inhibiting the production of unintended ligation products and increasing the capture efficiency of rare molecules with the
40 target regions while maintaining the uniformity of the product.

SUMMARY OF THE INVENTION

45 **[0004]** In view of the above-mentioned shortcomings of the prior art, the object of the present invention is to provide a gene target region enrichment method for solving the problems in the prior art.

[0005] To achieve the above objectives and other related objectives, the present invention provides a gene target region enrichment method, comprising:

- 50 (1) amplifying fragmented DNA comprising the target region via specific probes to provide captured-extension products, wherein the specific probe comprises a sequence complementary to the target region of the fragmented DNA, and the 3'terminal nucleotide of the specific probe is modified to prevent a ligation reaction at the 3' terminal of the specific probe;
- (2) linking the 3' terminal of the captured-extension product obtained in step (1) to linker DNA to obtain a ligation product.

55 **[0006]** Another aspect of the present invention provides a kit for enriching a fragmented DNA target region, which includes a specific probe and linker DNA(sometimes called "adapter DNA") suitable for the gene target region enrichment method.

BRIEF DESCRIPTION OF THE DRAWINGS

[0007]

FIG. 1 is a schematic flowchart of an enrichment process for a target region in an embodiment of the present invention.
 FIG. 2 is a schematic diagram of a library molecule constructed according to an embodiment of the present invention.
 FIG. 3 is a schematic diagram of the construction of the calibrator in an embodiment of the present invention.
 FIG. 4 is a schematic flowchart of a monospecific detection system in an embodiment of the present invention.

DETAILED DESCRIPTION OF THE EMBODIMENTS

[0008] The inventors of the present invention have provided a gene target region enrichment method after a lot of exploratory research. The gene target region enrichment method is easy to operate and has reliable results, especially for short-length nucleotide. Based on this, the present invention has been completed.

[0009] The first aspect of the present invention provides a gene target region enrichment method, comprising:

- (1) amplifying fragmented DNA comprising the target region via specific probes to provide captured-extension products, wherein the specific probe comprises a sequence complementary to the target region of the fragmented DNA, and the 3'terminal nucleotide of the specific probe is modified to prevent a ligation reaction at the 3' terminal of the specific probe;
- (2) linking the 3' terminal of the captured-extension product obtained in step (1) to linker DNA to obtain a ligation product.

[0010] The gene target region enrichment method provided by the present invention may include: amplifying the fragmented DNA comprising the target region by a specific probe to provide a captured-extension product. During the reaction, the specific probe can be extended by capturing the complementary sequence of the target region to obtain a Captured-Extension Product (CEP).

[0011] In the gene target region enrichment method, in the reaction of amplifying the fragmented DNA comprising the target region by a specific probe, the fragmented DNA comprising the target region may be one or more. Generally, there is a one-to-one correspondence between the specific probes and the fragmented DNA comprising the target region, that is, the number of specific probes in the reaction system may be one kind or multiple kinds. The extension described in this step refers to the entire pre-amplification step of the sample, including single or multiple rounds(cycles) of denaturation, annealing, and extension steps. In a preferred embodiment, in this step multiple cycles, such as 2-100, 2-10, 10-20, 20-30, 30-40, 40-60, 60-80, 80-100 cycles, are implements, in order to effectively increase the number of molecules containing the target area.

[0012] In the gene target region enrichment method, the fragmented DNA may be double-stranded DNA, single-stranded DNA, cDNA, etc. The cDNA may generally be obtained by reverse transcription of RNA. For double-stranded DNA, the specific probe may include a sequence complementary to the target region of one strand of the fragmented DNA. Therefore, the enrichment method of the present invention is also applicable to fragmented RNA. A person skilled in the art can perform subsequent operations through the enrichment method provided by the present invention after reverse-transcribing RNA into cDNA. The length of the fragmented DNA may be 25 to 200 bps, 25 to 40 bps, 40 to 60 bps, 60 to 80 bps, 80 to 100 bps, 100 to 120 bps, 120 to 140 bps, 140 to 160 bps, 160 to 180 bps, or 180 to 200 bps. In the gene target region enrichment method, the amplification system of step (1) may include specific probes, DNA polymerase and dNTP. The reaction of amplifying the fragmented DNA including target region via a specific probe can usually be carried out in the presence of DNA polymerase. After the probe with 3' end blocking-modified binds to the template under the action of high-fidelity polymerase, the blocking group is excised, and the probe is reactivated so that the target sequence can be extended effectively. The DNA polymerase may have 3'-5' exonuclease activity, so that the substituent group at the 3' end of the probe can be excised after binding to the template, so that the probe can extend along the template. The DNA polymerase is preferably a high-fidelity DNA polymerase so as to further improve the amplification efficiency and purity of the product. The DNA polymerase may also be a common DNA polymerase (i.e., it does not have 3'-5' exonuclease activity). The amplification system in the step (1) further includes an active substance. The active substance can be used to excise the 3'terminal modification group of the specific probe after binding to the target region. The active substance can also be combined with DNA polymerase (for example, common DNA polymerase) in the captured-extension system to improve the efficiency of the amplification system. The active substance is preferably a nuclease. The reaction of amplifying the fragmented DNA comprising the target region by a specific probe can generally be carried out in the presence of dNTP, which can usually be a dNTP coupled with a labeling molecule. The labeling molecule can include but not limited to biotin and the like. The dNTP may include but not limited to dCTP, dATP and the like. The dNTP may also be coupled with a labeling molecule, which may be biotin or the like, and the labeling molecule

may generally be used for purifying the captured-extension products.

[0013] In the gene target region enrichment method, the specific probe includes a sequence complementary to the target region of the fragmented DNA, thereby achieving specific amplification of the target region of the fragmented DNA. A person skilled in the art can select suitable target region of fragmented DNA, and design an appropriate complementary sequence according to the target region of fragmented DNA. The specific probe may be a specific probe whose 3' terminal nucleotide is modified, to prevent the 3' terminal of the specific probe from connecting with other groups, so as to avoid the free probe self-connecting or connecting other non-target molecules. For example, the non-target molecules may be linker DNA and the like. A person skilled in the art can select a suitable substituent group to modify the 3' terminal of the specific probe. For example, the modification group can replace the natural group (for example, hydroxyl group, etc.) on the 3' terminal nucleotide of the specific probe in order to prevent the ligation reaction at the 3' terminal of the specific probe. The modification group can usually be a blocking group. After the probe is combined with the target region on the template via the complementary sequence, the 3' terminal modification group can be excised by the enzyme, so that the probe is activated, and the target sequence can be effectively extended. The 3' terminal modification group of the specific probe may include, but not limited to, a hydrogen atom, a C3 Spacer group, a C6 Spacer group, a phosphate group (PO₄), an amino group (NH₂), etc.. The choice of different substituent groups has obvious differences in the capture effect of the probe. In a preferred embodiment of the present invention, substituting the 3' terminal hydroxyl group of the probe with C3 Spacer gives the best effect, which has obvious advantages compared with other substituent groups. The specific probe also includes a universal sequence, which can usually be recognized by the sequencing system, so that the ligation products provided subsequently can be sequenced by the sequencing system. For example, for the Ion Torrent sequencing system, the universal sequence can be corresponding PI sequence. In a preferred embodiment of the present invention, the base of the tail region of the 3' end of the specific probe may contain a mismatch, and the mismatched base may be the last base at the 3' end of the probe, or may be near the 3' end base. Mismatched base can be a single base, or multiple bases. Instead of affecting the specificity and binding efficiency of the probe, this mismatch will help to improve the cutting efficiency and fidelity of the high-fidelity DNA polymerase.

[0014] In the gene target region enrichment method, the step (1) may further include purification to the captured-extension product. A person skilled in the art may select a suitable method for purifying the captured-extension product. For example, the purification method for captured-extension product may include magnetic bead purification and the like. In a specific embodiment of the present invention, the captured-extension product can be purified by a labeling molecule. A solid phase purification method of magnetic beads coated with avidin or streptavidin can be used in the purification process.

[0015] The gene target region enrichment method provided by the present invention may further include: connecting the 3' end of the captured-extension product provided in step (1) to a linker DNA to provide a ligation product.

[0016] In the gene target region enrichment method, a ligation system that connects the 3' end of the captured-extension product to the linker DNA may include a single-stranded ligase that causes the 5' end of the linker DNA connected to the 3' end of the captured-extension product. The single-stranded ligase is preferably T4 RNA ligase or thermostable RNA ligase or the like.

[0017] In the gene target region enrichment method, the linker DNA may be modified at the 5' terminal nucleotide and has a single-stranded structure at the reaction temperature of step (2), so that under the catalysis of the chain ligase, a covalent bond is formed between 3' end hydroxyl group of the captured-extension product and the 5' end modification group of the linker DNA and the ligation product is obtained. In a specific embodiment of the present invention, the 5' terminal nucleotide of the linker DNA (for example, the 5-position hydroxyl group of the 5' terminal nucleotide) is substituted with a phosphate group, thereby forming a phosphorylation modification. Under the catalysis of the single-stranded ligase, the 3' end hydroxyl group of the captured-extension product and the phosphorylated 5' end of the linker DNA form a covalent bond to obtain the ligation product. In another specific embodiment of the present invention, the 5' terminal nucleotide of the linker DNA (for example, specifically the 5-position hydroxyl group of the 5' terminal nucleotide) is substituted with an adenosine group, thereby forming an adenylation modification. Under the catalysis of 5'App DNA/RNA thermostable ligase, the 5' end of the linker DNA can be ligated with the 3' end of the captured-extension product. In another specific embodiment of the present invention, the 5' terminal nucleotide of the linker DNA (for example, specifically the 5-position hydroxyl group of the 5' terminal nucleotide) is substituted with a phosphate group, thereby forming a phosphorylation modification. Under the catalysis of thermostable RNA ligase, the 5' terminal of the linker DNA can be ligated with the 3' end of the captured-extension product. The linker DNA may also be a partially double-stranded structure with a sticky end at the 5' end region, and the sticky end at the 5' end region has single-stranded property, so that the 5' end can be modified by the method described above. Under the catalysis of a suitable single-stranded ligase, the 5' end can be linked to the captured-extension product. The linker DNA may also include one or more of universal sequence, sample tag sequence, molecular tag sequence, etc. These sequences can usually be recognized by a sequencing system, so that the obtained ligation product can be sequenced. For example, for the Ion Torrent sequencing system, the universal sequence may be sequence A, the sample tag sequence may be barcode, the molecular tag sequence

corresponding thereto, and the like. Introducing barcode is helpful to distinguish samples from different sources in subsequent bioinformatics analysis, so that multiple samples can be sequenced simultaneously in a single reaction to increase throughput. The introduction of molecular tags facilitates the identification of different molecules in the subsequent bioinformatics analysis, so as to further identify the mutations generated in the subsequent amplification. The application of these sequences that can be recognized by the sequencing system enables the library of the present invention to be sequenced through the high-throughput sequencing platform to provide the information required for various subsequent research and clinical applications.

[0018] In a preferred embodiment of the present invention, the 3' end hydroxyl group of the linker DNA is also substituted by other groups to prevent the ligation reaction of the 3' end of the linker DNA to avoid self-ligation or ligating with non-target molecule, such as the free probe from step (1).

[0019] In the gene target region enrichment method, the step (2) may further include purification to the ligation product. A person skilled in the art may select a suitable method to purify the ligation product. For example, the purification method to the ligation product may include but not limited to silica gel column purification, heat treatment, and the like.

[0020] The gene target region enrichment method provided by the present invention can be used for gene detection. Methods for further gene detection to the amplified ligation products are known to a person skilled in the art. In a specific embodiment of the present invention, the target region enrichment method can be applied to high-throughput sequencing. In another specific embodiment of the present invention, the method of enriching the target region of the present invention can be applied to the detection of gene sequences. For example, the target region includes: a site where the sequence mutated, and more specifically, a single base mutation site area, base deletion site area, base insertion site area and fusion mutation site area, etc. In another specific embodiment of the present invention, the method of enriching the target region of the present invention can be applied to the detection of the state of gene modification. For example, the target region includes a region where methylation sites may exist, or, the DNA fragment has been treated with bisulfite.

[0021] The gene target region enrichment method provided by the present invention may further include: (3) amplifying the ligation product provided in the step (2). Generally, the ligation product can be PCR amplified via DNA polymerase and PCR amplification primers. By amplifying the ligation product, the product containing the DNA of the target region can be further enriched. A person skilled in the art may select a suitable method and system to amplify the ligation product provided in step (2), for example, the PCR amplification primer has a sequence matching the universal sequence of the specific probe and/or the universal sequence of the linker DNA. Specifically, the PCR amplification primer has a sequence that is at least partially complementary to the universal sequence of the specific probe and the universal sequence of the linker DNA.

[0022] The gene target region enrichment method provided by the present invention may further include: (4) sequencing the amplified ligation products to provide sequencing results of the target region. After PCR amplification of the ligation product, the amplification product can be sequenced by identifying the universal sequence in the amplification product to obtain the sequencing result of the target region. A person skilled in the art may select a suitable method and system to sequence the amplified ligation product, for example, sequencing the P1 sequence and the A sequence based on the Ion Torrent platform to obtain the sequencing result of the target region. It should be noted that the enrichment method of the present invention can make the ligation product obtained in step (2) contain any universal sequence for sequencing, such as universal sequence of Ion Torrent sequencing platform, Illumina sequencing platform, or other sequencing platforms.

[0023] The gene target region enrichment method provided by the present invention may further include: (5) detecting the ligation product provided in step (2) by the detection primer 1, the detection primer 2 and the probe 3, enabling the rapid detection of the target region by means of PCR instead of next-generation sequencing. At least one of the detection primer 1, the detection primer 2 and the probe 3 contains a gene-specific sequence, that is, for different gene target regions, the combination of the three primer/probe's specific sequences can be monospecific, bispecific or trispecific. In an embodiment of the present invention, the detection primer 1 includes a gene-specific sequence, the detection primer 2 and the probe 3 include universal sequence; or the detection primer 1 and the detection primer 2 include a gene-specific sequence; or the detection primer 1 and the probe 3 contains a gene-specific sequence; or all of detection primer 1, detection primer 2, and probe 3 contain a gene-specific sequence. In some embodiments, the probe 3 may further include a labeling molecule, such as a fluorescent molecule, and the sequence of probe 3 is not complementary to that of detection primer 1 or 2.

[0024] The second aspect of the present invention provides a kit for enriching target region of fragmented DNA, comprising specific probe and linker DNA suitable for the gene target region enrichment method provided in the first aspect of the present invention. The structure of the specific probe and the linker DNA has been described in detail in the first aspect of the present invention, and will be omitted here.

[0025] The kit provided by the present invention may further include one or more of the following components: RNA ligase, dNTP coupled with a labeling molecule, DNA polymerase, nuclease, etc. Among them, the RNA ligase can be a thermostable RNA ligase, T4 RNA ligase, or 5'App DNA/RNA thermostable ligase, and the like. dNTP can be dCTP or dATP. The DNA polymerase may be a DNA polymerase having 3'-5' exonuclease activity, preferably, the DNA

polymerase is a high-fidelity polymerase.

[0026] The kit provided by the present invention may also include forward primers and reverse primers used for PCR amplification, which generally matches the universal sequence of the specific probe and the universal sequence of the linker DNA. In particular, the primer may have a sequence at least partially complementary to the universal sequence of the specific probe and the universal sequence of the linker DNA.

[0027] The kit provided by the present invention may further include a detection primer 1, a detection primer 2 and a probe 3 for PCR detection, at least one of them contains a gene-specific sequence, in particular, combination of monospecific, bispecific or trispecific primer/probe. Preferably, only detection primer 1 contains a gene-specific sequence, refer to Figure 4 for the detection process.

[0028] In a preferred embodiment of the present invention, after library construction by the method or kit of the present invention, as shown in Figure 2, library molecules (for example, the ligation product provided in step (2), the structure is as shown in FIG. 2) include the following sequences in order: 5' end sequencing universal sequence, gene specific probe sequence, enriched target region sequence, adapter(linker) sequence, barcode, molecular tag sequence, 3' end sequencing universal sequence. The enriched target region contains the sequence information of the sample DNA before enrichment. The characteristics of these sequences are as follows: the position of the 5' end on the genome is fixed and determined by the specific probe, while the position of the 3' end is not fixed, determined by the initial DNA fragmentation status of the library. Therefore, in the data analysis after enrichment, the position of the 3' end of the sequence on the genome can also serve as a molecular tag. Molecular tags can be used to distinguish different molecules and improve detection sensitivity and accuracy.

[0029] The beneficial effects of the present invention are as follows:

Firstly, before the ligation, the target regions of all fragmented DNA samples are preamplified by means of captured-extension, to avoid loss or missed detection of original target molecules(especially small fragments and rare molecules) caused by insufficient ligation efficiency of ligase during the ligation stage. The extension reaction in the pre-amplification stage is linear amplification and has no preference of PCR amplification, so, it will not accumulate errors introduced by PCR amplification. Compared with the conventional PCR library constructing technology, the product is more uniform and the subsequent sequencing results are more accurate.

Secondly, in the pre-amplification stage, only a single-stranded probe with a length of about 30 bp is necessary to be designed for each target gene, avoiding the difficulty of designing double-ended primers for short fragments such as cfDNA and shorter fragments such as ctDNA. Thus, not only the success rate of library construction but also the convenience of library construction is improved. Blocking the 3'end of the probe can block the non-intended ligation except the probe and the template, effectively reducing the background noise caused by the free probe, supplemented by the dNTP coupled to the labeling molecule and its purification system, the purity of the target product can be further improved and the sample DNA molecules can reach the theoretical maximum conversion rate.

Thirdly, after the 5' end of the linker DNA is modified, the linker DNA can be well connected to the 3' end of the extension product under the catalysis of single-stranded ligase and will not be connected to the 3' end of the block-modified specific probe. The 3' end of the linker is blocked to avoid self-ligation or mis-ligation of the linker and increase the proportion of the target ligation product in the final product.

Finally, in the process of enriching the target region, the universal sequence for sequencing, the sample tag sequence and the molecular tag sequence are designed to be contained in the sequences of the probe and the linker, enabling direct sequencing and data analysis after library construction; or, rapid PCR detection of the target region can be achieved using single, dual or trispecific primer/probe system.

[0030] In summary, the enrichment method and kit of the present invention are simple in operation and produces reliable results. The use of DNA with a fragment length of less than 200bp can minimize the loss of original molecules, especially rare molecules, and enrich the target most efficiently. In addition, the enrichment method and kit have extremely high fidelity, and the base error generated during the sequencing process can be effectively corrected through subsequent bioinformatics analysis, and the highest sequencing accuracy rate in theory can be achieved.

[0031] The following describes the embodiments of the present invention through specific examples. A person skilled in the art can easily understand other advantages and effects of the present invention from the contents disclosed in this specification. The present invention can also be implemented or applied through different specific embodiments. Based on different viewpoints and applications, the details in this specification can also be modified or changed without departing from the spirit of the present invention.

[0032] Before further describing the specific embodiments of the present invention, it should be noted that the protection scope of the present invention is not limited to the specific embodiments described below. It should also be noted that the terms used in the examples of the present invention are intended to describe specific embodiments of the present invention, and are not intended to limit the scope of protection of the present invention. In the description and claims of the present invention, unless clearly stated otherwise, the singular forms "a", "an", "one" and "this" include plural forms.

[0033] When the numerical ranges are recorded in the embodiments, it should be noted that unless the invention indicates otherwise, the two endpoints of each numerical range and any one value between the two endpoints can be selected. Unless otherwise defined, all technical and scientific terms used in the present invention have the same meaning as commonly understood by a person skilled in the art. In addition to the specific methods, equipment, and materials used in the embodiments, the methods, equipment, and materials described in the embodiments of the present invention can also be used according to the master of the prior art and the description of the present invention by a person skilled in the art. Similar or equivalent methods, devices and materials in prior art can also be used to implement the present invention.

[0034] Unless otherwise stated, the experimental methods, detection methods, and preparation methods disclosed in the present invention use conventional technologies in molecular biology, biochemistry, chromatin structure and analysis, analytical chemistry, cell culture, and reorganization DNA technology and related fields. These technologies have been well described in the existing literature. For details, see Sambrook et al., MOLECULAR CLONING : A LABORATORY MANUAL , Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol.304, Chromatin (P.M. Wasarman and A.P. Wolffe, eds.) , Academic Press, San Diego, 1999; and METHODS IN MOLECULAR BIOLOGY, Vol. 119, Chromatin Protocols (P.B. Becker, ed.) Humana Press, Totowa, 1999, etc.

Example 1

[0035] The oligonucleotide sequences used in Example 1 are shown in table 1:

Table 1

Name	SEQ ID NO.	Sequence (5'→3')
Probe 1	1	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-OH
Probe 2	2	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-C3 Spacer
Probe 2p	3	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-PO ₄
Probe 2c	4	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-C6 Spacer
Probe 2n	5	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-NH ₂
Probe 2d	6	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAA-DDC
Probe 3	7	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGGTGAT-C3 Spacer
Probe 4	8	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCCTGGCAGCCAGGAACGT ACTGrGTGAT-C3 Spacer
ABar-X1	9	PO ₄ - GGATCCNNNNNNNCAGCTTGGACTGAGTCGGAGACACGCAGGGATGA GATGG

(continued)

Name	SEQ ID NO.	Sequence (5'→3')
ABar-X2	10	PO ₄ - GGATCCNNNNNNNGTTCTCCTTACTGAGTCGGAGACACGCAGGGATGA GATGG-C3 Spacer
ABar-X3	11	5' App- GGATCCNNNNNNNTTACCTTAGCTGAGTCGGAGACACGCAGGGATGAG ATGG-C3 Spacer
EF-1	12	GATCACAGATTTTGGGC
ER-1	13	TTGCCTCCTTCTGC
EM-1	14	AACTGCTGGGTGCGGA
AF	15	GTCTCAGCCTCTCTATGGGCAGTCGGTGAT
AR	16	CCATCTCATCCCTGCGTGTCTCCGACTCAG

[0036] The probe in this embodiment contains a universal sequence for sequencing (the bases marked in *italics*) and here contains the PI sequence in the Ion Torrent sequencing system. The specific sequences of all probes in the table are directed to the L858R mutation of exon 21 of the EGFR gene. Among them, the probe 1 (SEQ ID NO. 1) has no modification at the 3' terminal, and 3' terminal hydroxyl groups of probe 2, 3 and 4 (SEQ ID NO. 2, SEQ ID NO. 7, SEQ ID NO. 8) are substituted with C3 Spacer, 3' terminal hydroxyl group of probe 2p (SEQ ID NO. 3) is substituted with a phosphate group, 3' terminal hydroxyl group of probe 2c (SEQ ID NO. 4) is substituted with C6 Spacer, 3' terminal hydroxyl group of probe 2n (SEQ ID NO. 5) was substituted with a phosphate group, and 3' terminal hydroxyl group of probe 2d (SEQ ID NO. 6) is substituted with dideoxycytosine (DDC). Among them, the last base of probe 3 (SEQ ID NO. 7) is a mismatched base, and the G near the 3' end in probe 4 (SEQ ID NO. 8) is substituted with the RNA base rG.

[0037] The linker ABar-X1/2/3 (SEQ ID NO. 9, SEQ ID NO. 10, SEQ ID NO. 11) contains universal sequence for sequencing, the bases marked in *italics*, here is Ion Torrent sequence A in the sequencing system. The underlined part of the sequence is barcode, which can be substituted with different barcodes in different samples of the present invention, and the changes in this part will be omitted in the present invention. The "N" is a molecular tag sequence, wherein NNNNNN is a random sequence used to label different extension product molecules in the same sample. The 5' terminal of the ABar-X1 and ABar-X2 (SEQ ID NO. 9, SEQ ID NO. 10) is phosphorylated modification. The 5' end of the ABar-X3 (SEQ ID NO. 11) is adenylation modification. The hydroxyl groups at the 3' end of ABar-X2 and ABar-X3 (SEQ ID NO. 10, SEQ ID NO. 11) are substituted with C3 Spacer.

[0038] EF-1 and ER-1 (SEQ ID NO. 12, SEQ ID NO. 13) are the front and rear primers used for PCR detection of EGFR target sequence, wherein EM-1 (SEQ ID NO. 14) is MGB probe. AF and AR (SEQ ID NO. 15, SEQ ID NO. 16) are the front and rear primers used to amplify the library, respectively, and to match the universal sequence in the library, namely the PI sequence in the probe and the A sequence in the linker.

Main reagents and materials

[0039] Plasma free DNA extraction kit was purchased from Qiagen; Cell DNA extraction kit and purified single-stranded DNA kit were RNA Clean Kit purchased from Tiangen Biotech; Thermostable RNA ligase was purchased from Epicentre, USA; High-fidelity DNA polymerase reaction system, Biotin-dCTP and MyOne Strptavidin C1 reagent for purification were purchased from Invitrogen. DNA polymerase, high-fidelity DNA polymerase and PCR reaction kit were purchased from TOYOBO. RNase H₂ thermostable nuclease was purchased from IDT. Q5 high-fidelity DNA polymerase, T4 RNA ligase and 5'App DNA/RNA thermostable ligase were purchased from NEB. Agencourt AMPure magnetic beads were purchased from Beckman. The calibrators for quantitative detection in the relevant process were constructed according to the conventional methods of molecular cloning, and the detailed composition is shown in Figure 3.

Step 1. Sample preparation

[0040] Extract cfDNA in healthy human plasma samples and DNA in NCI-H1975 cell line (this cell line is mutation of

EP 3 907 296 A1

exon 21 of EGFR gene L858R) with extraction kit, and the DNA samples are quantified by Qubit Fluorometer quantifier. The DNA of the NCI-H1975 cell line was broken into fragments of about 160 bp, and mixed into the cfDNA samples of healthy human at the proportions of 10%, 1%, 0.1%, and 0.01%, and the cfDNA of healthy human was used as a blank control (QC).

Step 2. Captured-extension

[0041] Prepare the captured-extension reaction system according to table 2, table 3, and table 4, respectively.

Table 2

Compositions	Volume (μ L)	Final Concentration
2X high-fidelity DNA polymerase PCR Master Mix (with 3'-5' exonuclease activity); or, 2X DNA polymerase PCR Master Mix (without 3'-5' exonuclease activity)	25	1×
Specific probe (10 μ M)	0.5	100 nM
cfDNA samples of all types	20	20000 copies
H ₂ O	4.5	/
Total	50	/

Table 3

Compositions	Volume (μ L)	Final Concentration
2X DNA polymerase PCR Master Mix (without 3'-5' exonuclease activity)	25	1×
RNAse H ₂ thermostable nuclease	0.5	/
Specific probe (10 μ M)	0.5	100 nM
cfDNA samples of all types	20	20000 copies
H ₂ O	4	/
Total	50	/

Table 4

Compositions	Volume (μ L)	Final Concentration
2X high-fidelity DNA polymerase PCR Master Mix (with 3'-5' exonuclease activity)	25	1×
Specific probe (10 μ M)	0.5	100 nM
Biotin-dCTP (0.4 mM)	2.5	20 μ M
cfDNA samples of all types	20	20000 copies
H ₂ O	2	/
Total	50	/

[0042] The PCR procedure of captured-extension is shown in table 5:

Table 5

Cycles	Temperature	Time
1 cycle	98°C	60 s

(continued)

Cycles	Temperature	Time
1/20/40 cycles	98°C	10 s
	67°C	20 s
	72°C	20 s
1 cycle	72°C	5 min

[0043] After completion of the captured-extension process, the extension product was purified by silica gel column using RNA Clean kit and eluted with 60 μ L of the elution buffer. (The silica gel column purification mentioned hereafter in the present invention is the same). The extension product was further purified with the magnetic beads coated with streptavidin after the silica gel column purification, and finally all the products are dissolved in 60 μ L elution buffer.

Step 3. Extension efficiency detection

[0044] Prepare PCR detection system for extension products according to table 6.

Table 6

Compositions	Volume (μ L)	Final Concentration
2 $\times$ Taqman Mix	25	1 $\times$
ER-1 (10 μ M)	1.5	300 nM
AF (10 μ M)	1.5	300 nM
EM-1 (10 μ M)	0.5	100 nM
H ₂ O	17.5	/
Extension product/Calibrator	4	/
Total Volume	50	/

[0045] The PCR procedure for detecting the extension efficiency is shown in table 7.

Table 7

Cycles	Temperature	Time
1 cycle	95°C	3 min
45 cycles	95°C	10 s
	56°C	25 s
	72°C	20 s

Step 4. Single-stranded ligation

[0046] According to the results of step 3, select part of the extension product obtained in step 2 to prepare the ligation reaction system according to table 8, table 9, and table 10, respectively.

EP 3 907 296 A1

Table 8

Compositions	Volume (μL)	Final Concentration	Ligation Conditions
CircLigase II 10X Reaction Buffer	4	1×	60°C 1 hour
50 mM MnCl_2	2	2.5 mM	
5 M Betaine	8	1M	
Thermostable RNA ligase (100 U/ μL)	2	5 U/ μL	
ABar-X1/ABar-X2 (200 nM)	2	10 nM	
Extension product	8		
50% PEG	14	17.5%	
Total	40	/	

Table 9

Compositions	Volume (μL)	Final Concentration	Ligation Conditions
10X Buffer for ligation	4	1×	16°C 4 hours
T4 RNA ligase	2	5 U/ μL	
Linker ABar-X2(200 nM)	2	10 nM	
Extension Product	8	/	
50% PEG	14	17.5%	
Total	40	/	

Table 10

Compositions	Volume (μL)	Final Concentration	Ligation Conditions
10X NE Buffer 1	4	1×	65°C 1 hour
50 mM MnCl_2	2	2.5 mM	
H_2O	8	/	
5'App DNA/RNA thermostable ligase (100 U/ μL)	2	5 U/ μL	
Linker ABar-X3 (200 nM)	2	10 nM	
Extension Product	8		
50% PEG	14	17.5%	
Total	40	/	

[0047] Treatment process after obtaining the ligation product is shown in table 11.

Table 11

Treatment 1	Treatment 2	Treatment 3
None (as control)	Incubating for 10 min at 95°C	Purification using silica gel column

Step 5. Test of the ligation efficiency

[0048] Prepare the PCR system for testing the ligation efficiency of step 4 according to table 12.

EP 3 907 296 A1

Table 12

Reagent	Dosage (μL)	Final Concentration
2X Taqman Mix	12.5	1×
AF (10 μM)	0.75	300 nM
AR (10 μM)	0.75	300 nM
EM-1(10 μM)	0.25	100 nM
H ₂ O	8.75	/
Calibrator /Ligation Product	2	/
Total	25	/

[0049] PCR reaction conditions for testing the ligation efficiency is shown in table 13.

Table 13.

Cycles	Temperature	Time
1 cycle	95°C	4 min
45 cycles	95°C	10 s
	60°C	30 s
	72°C	20 s

Step 6. Amplification of the Library

[0050] Amplify the ligation product obtained in step 4 with PCR according to the reaction conditions and procedure as shown in table 14.

Table 14

Compositions		Volume (μL)
Ligation product		30
5× Q5 Buffer		10
dNTP(10 mM each)		1
Q5 high-fidelity DNA polymerase (2 U/ μL)		1
AF (10 μM)		1.5
AR (10 μM)		1.5
ddH ₂ O		5
Total Volume		50
Temperature(°C)	Time(s)	Cycles
98	30	1
98	10	15
72	30	
72	300	1

[0051] Purify the amplified library with 80 μL of Agencourt AMPure magnetic beads, and dissolve the purified amplified product in 30 μL of elution buffer. At this point, the library construction is completed, and the library is ready for sequencing.

Step 7. Sequencing and data analysis

[0052] Sequencing the prepared library on the Ion Proton sequencer. Related operations include water-in-oil PCR, library enrichment, chip loading, and on-board sequencing. For detailed operation procedures, see Ion PITM Hi-Q™ OT2 200 Kit manual and Ion PITM Hi-Q™ Sequencing 200 Kit manual.

Experimental results and analysis

[0053] The calculation methods for captured-extension multiple(magnification), capture efficiency and ligation efficiency in the results are as follows:

Captured-extension multiple (magnification) = EGFR gene output copies after capture / EGFR gene input copies before capture

Capture efficiency = (Captured-extension multiple-1)/Captured-extension cycles×100%

Ligation efficiency = (Output copies of ligation product / Input copies of extension product) × 100%

[0054] The analysis of all process data (such as capture efficiency, ligation efficiency) in this embodiment comes from QC samples.

[0055] Results 1. The effect of different types of polymerases on the captured-extension results of different probes.

Table 15. Testing results of different types of polymerases for captured-extension (QC sample)

Probe	Types of DNA polymerases	Nuclease	Cycles	Copies of <i>EGFR</i> output after captured-extension	Copies of <i>EGFR</i> input before captured-extension	Captured-extension multiple	Capture Efficiency%
Probe 1	High-fidelity polymerase	N	40	482000	20000	24.1	57.75%
Probe 2	High-fidelity polymerase	N	40	566000	20000	28.3	68.25%
Probe 2p	High-fidelity polymerase	N	40	98400	20000	4.92	9.80%
Probe 2c	High-fidelity polymerase	N	40	428000	20000	21.4	51%
Probe 2n	High-fidelity polymerase	N	40	64800	20000	3.24	5.6%
Probe 2d	High-fidelity polymerase	N	40	71200	20000	3.56	6.4%
Probe 3	High-fidelity polymerase	N	40	524000	20000	26.2	63.00%
Probe 4	High-fidelity polymerase	N	40	482000	20000	24.1	57.75%
Probe 1	Common polymerase	N	40	512000	20000	25.6	61.50%
Probe 2	Common polymerase	N	40	24000	20000	1.2	0.50%

(continued)

Probe	Types of DNA polymerases	Nuclease	Cycles	Copies of <i>EGFR</i> output after captured-extension	Copies of <i>EGFR</i> input before captured-extension	Captured-extension multiple	Capture Efficiency%
Probe 2p	Common polymerase	N	40	21000	20000	1.05	0.12%
Probe 2c	High-fidelity polymerase	N	40	21000	20000	1.1	0.25%
Probe 2n	High-fidelity polymerase	N	40	23000	20000	1.15	0.38%
Probe 2d	High-fidelity polymerase	N	40	24000	20000	1.20	0.5%
Probe 3	Common polymerase	N	40	22000	20000	1.1	0.25%
Probe 4	Common polymerase	N	40	26000	20000	1.3	0.75%
Probe 1	Common polymerase	Y	40	496000	20000	24.8	59.50%
Probe 2	Common polymerase	Y	40	26000	20000	1.3	0.75%
Probe 3	Common polymerase	Y	40	22000	20000	1.1	0.25%
Probe 4	Common polymerase	Y	40	506000	20000	25.3	60.75%

[0056] The captured-extension reaction system involved in table 15 is shown in table 2 and table 3, wherein, the high-fidelity polymerase is 2X high-fidelity DNA polymerase PCR Master Mix (DNA polymerase has 3'-5' exonuclease activity), the common polymerase is 2X DNA polymerase PCR Master Mix (DNA polymerase does not have 3'-5' exonuclease activity), the cfDNA sample used is QC sample. Please refer to step 3 for the method of extension efficiency test.

[0057] It can be seen from the test results: 1) After the 3' end blocking-modified probe binds to the template under the action of high-fidelity polymerase, the blocking group is excised and the probe is activated, so that the target sequence can be extended effectively; 2) Combination of thermostable nuclease and common polymerase play an activation effect to the probe with modified RNA, enhancing the applicability of the enrichment method of the present invention; 3) Selection of different substitution groups has obvious differences in the capture effect of the probe. Substituting the 3'terminal hydroxyl group of the probe with C3 Spacer gives the best effect, which has obvious advantages compared with other substituent groups. Therefore, the data from result 2 in this example mainly refers to the result data obtained for the probe in which the 3' terminal hydroxyl group is substituted by C3 Spacer.

[0058] Result 2. The effect of different extension cycles on probe extension results.

Table 16. Test results of different extension cycles (QC sample)

Probe	Types of DNA polymerases	Nuclease	Cycles	Copies of <i>EGFR</i> output after captured-extension	Copies of <i>EGFR</i> input before captured-extension	Captured-extension Magnification	Capture Efficiency %
Probe 1	High-fidelity polymerase	N	40	482000	20000	24.1	57.75%
Probe 2	High-fidelity polymerase	N	40	566000	20000	28.3	68.25%

(continued)

Probe	Types of DNA polymerases	Nuclease	Cycles	Copies of <i>EGFR</i> output after captured-extension	Copies of <i>EGFR</i> input before captured-extension	Captured-extension Magnification	Capture Efficiency %
Probe 3	High-fidelity polymerase	N	40	524000	20000	26.2	63.00%
Probe 4	High-fidelity polymerase	N	40	482000	20000	24.1	57.75%
Probe 1	High-fidelity polymerase	N	20	276000	20000	13.8	64.00%
Probe 2	High-fidelity polymerase	N	20	244000	20000	12.2	56.00%
Probe 3	High-fidelity polymerase	N	20	234000	20000	11.7	53.50%
Probe 4	High-fidelity polymerase	N	20	248000	20000	12.4	57.00%
Probe 1	High-fidelity polymerase	N	1	34800	20000	1.74	74.00%
Probe 2	High-fidelity polymerase	N	1	34400	20000	1.72	72.00%
Probe 3	High-fidelity polymerase	N	1	36400	20000	1.82	82.00%
Probe 4	High-fidelity polymerase	N	1	31400	20000	1.57	57.00%

[0059] The captured-extension system referred to in Table 16 is disclosed in Table 2. The cfDNA sample used in the result is the QC sample. The method for testing the extension efficiency is shown in step 3.

[0060] The results showed that the Captured-extension magnification can be effectively increased by increasing the extension cycles.

Result 3. Effect of dNTP coupling with labeling molecule on the captured-extension performance

[0061]

Table 17. Test results of captured-extension system with Biotin-dCTP mixed(QC sample)

Probe	Types of DNA polymerases	Purification system with Biotin-dCTP mixed	Cycles	Copies of <i>EGFR</i> output with biotin labeled	Copies of <i>EGFR</i> input before capture	Captured extension Magnification	Capture Efficiency%
Probe 1	High-fidelity polymerase	N	40	0	20000	0	/
Probe 2	High-fidelity polymerase	N	40	0	20000	0	/
Probe 3	High-fidelity polymerase	N	40	0	20000	0	/
Probe 4	High-fidelity polymerase	N	40	0	20000	0	/

(continued)

Probe	Types of DNA polymerases	Purification system with Biotin-dCTP mixed	Cycles	Copies of <i>EGFR</i> output with biotin labeled	Copies of <i>EGFR</i> input before capture	Captured extension Magnification	Capture Efficiency%
Probe 1	High-fidelity polymerase	Y	40	212000	20000	10.6	24.00%
Probe 2	High-fidelity polymerase	Y	40	183000	20000	9.15	20.38%
Probe 3	High-fidelity polymerase	Y	40	198000	20000	9.9	22.25%
Probe 4	High-fidelity polymerase	Y	40	187000	20000	9.35	20.88%

[0062] The captured-extension system referred to in Table 17 is disclosed in Table 2 and 4. The cfDNA sample used in the result is the QC sample. The method for testing the extension efficiency is shown in step 3.

[0063] The results showed that the extension products can be labeled effectively and the Capture Efficiency can be effectively increased by mixing dCTP coupled with biotin.

Result 4. Effect of dNTP coupling with labeling molecule on the ligation efficiency

[0064]

Table 18. Test result of the ligation reaction after mixing Biotin-dCTP into the captured-extension system (QC)

Probe	Purification system with Biotin-dCTP mixed	Linker	Ligase	Purification method after ligation	Copies of extension input	Copies of ligation output	Ligation Efficiency
Probe 1	N	ABar-X2	Thermostable RNA ligase	Silica gel column	77120	84	0.11%
Probe 2	N	ABar-X2	Thermostable RNA ligase	Silica gel column	90560	1452	1.60%
Probe 3	N	ABar-X2	Thermostable RNA ligase	Silica gel column	83840	1878	2.24%
Probe 1	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	28267	328	1.16%
Probe 2	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	24400	13257	54.33%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	26400	17328	65.64%

[0065] The ligation reaction system referred to in Table 18 is shown in Table 8, wherein the extension product is prepared according to the system of Table 2 or Table 4, wherein high-fidelity polymerase is used in each system and the high-fidelity polymerase is 2X high-fidelity DNA polymerase PCR Master Mix (with a 3'-5' exonuclease activity). The cfDNA sample used is QC sample. The method for testing the ligation efficiency is shown in step 5.

[0066] The results showed that probes with 3' terminal hydroxyl group substituted by C3 Spacer in combination with using the dCTP coupled with biotin and a purification system targeting the biotin gives the best results; while in systems comprising probes without substituent group or dNTP without biotin, almost unable to ligate(connect).

Result 5. Effect of probes with different 5' terminal substitution and the use of different ligases on the ligation efficiency.

[0067]

Table 19. Test results of probes with different 5' terminal substitution and different ligases in the ligation system (QC sample)

Probe	Purification system with Biotin-dCTP mixed	Linker	Ligase	Purification method after ligation	Copies of extension input	Copies of ligation output	Ligation Efficiency
Probe 2	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	24400	13257	54.33%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	26400	17328	65.64%
Probe 2	Y	ABar-X2	T4 RNA	Silica gel column	24400	2264	9.28%
Probe 3	Y	ABar-X2	T4 RNA	Silica gel column	26400	2828	10.71%
Probe 2	Y	ABar-X3	5'AppDNA/RNA ligase	Silica gel column	24400	2073	8.50%
Probe 3	Y	ABar-X3	5'AppDNA/RNA ligase	Silica gel column	26400	2154	8.16%

[0068] The ligation system referred to in Table 19 is shown in Table 8, 9 and 10, wherein the extension product is prepared according to the system of Table 4. The cfDNA sample used is the QC sample. The method for testing the ligation efficiency is shown in step 5.

[0069] The results showed that probes with phosphorylated 5' terminal in combination with a thermostable RNA ligase can significantly improve the ligation efficiency.

Result 6. Effect of linker DNA with different 3' terminal substitution on the ligation efficiency.

[0070]

Table 20. Test results of ligation efficiency using linker DNA with different 3' terminal substitution in the system (QC sample)

Probe	Purification system with Biotin-dCTP mixed	Linker	Ligase	Purification method after ligation	Copies of extension input	Copies of ligation output	Ligation Efficiency
Probe 2	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	24400	13257	54.33%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	26400	17328	65.64%
Probe 2	Y	ABar-X1	Thermostable RNA ligase	Silica gel column	24400	4321	17.71%
Probe 3	Y	ABar-X1	Thermostable RNA ligase	Silica gel column	26400	3875	14.68%

[0071] The ligation system referred to in Table 20 is shown in Table 8, wherein the extension product is prepared according to Table 4. The cfDNA sample used is the QC sample. The method for testing the ligation efficiency is shown in step 5.

[0072] The results showed that linkers with blocked 3' terminal significantly improves the ligation efficiency by reducing self-ligation and mis-ligation.

Result 7. Effect of different purification methods after ligation on the proportion of the target product.

[0073]

Table 21. Results of the proportion of the target product by using different purification methods after ligation (QC sample)

Probe	Purification system with Biotin-dCTP mixed	Linker	Ligase	Purification method after ligation	Copies of extension input	Copies of ligation output	proportion of the target product
Probe 2	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	24400	13257	54.33%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	Silica gel column	26400	17328	65.64%
Probe 2	Y	ABar-X2	Thermostable RNA ligase	Heating	24400	10567	43.31%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	Heating	26400	12385	46.91%
Probe 2	Y	ABar-X2	Thermostable RNA ligase	None	24400	2324	9.52%
Probe 3	Y	ABar-X2	Thermostable RNA ligase	None	26400	2875	10.89%

[0074] The ligation reaction system referred to in Table 21 is shown in Table 8, wherein the extension product is prepared according to Table 4. The cfDNA sample used is the QC sample. The specific parameters of purification method used after ligation is shown in Table 11. The method for testing the ligation efficiency is shown in step 5.

[0075] The results showed that after the ligation reaction, either the silica gel column purification or heating purification can effectively remove or reduce the amount of residual thermostable RNA ligase in the ligation products, thereby reducing the background noise in ligation products and increasing the detection rate of proportion of the target product.

Result 8. Data of the samples under different reaction conditions.

[0076] The reaction conditions for the sequencing analysis are shown in Table 22. The sequencing method is disclosed in step 7. The results of the sequencing analysis corresponding to the conditions as stated in Table 22 are shown in Table 23. The detection results of the cfDNA in which the mutant DNA is mixed with different ratios are shown in Table 24.

Table 22. Reaction conditions of the sequencing analysis

Condition Number	Probe	Type of DNA polymerase	Nuclease	Cycles	Purification system with Biotin-dCTP mixed	Linker	Ligase	Process after ligation
1	Probe 1	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	Silica gel column
2	Probe 2	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	Silica gel column
3	Probe 2p	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	Silica gel column
4	Probe 3	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	Silica gel column
5	Probe 4	Common polymerase	Y	40	Y	ABar-X2	Thermostable RNA ligase	Silica gel column
6	Probe 3	High-fidelity polymerase	N	40	N	ABar-X2	Thermostable RNA ligase	Silica gel column
7	Probe 3	High-fidelity polymerase	N	40	Y	ABar-X1	Thermostable RNA ligase	Silica gel column
8	Probe 3	High-fidelity polymerase	N	40	Y	ABar-X2	T4 RNA	Silica gel column
9	Probe 3	High-fidelity polymerase	N	40	Y	ABar-X3	5'AppDNA/RNA ligase	Silica gel column
10	Probe 3	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	Heating
11	Probe3	High-fidelity polymerase	N	40	Y	ABar-X2	Thermostable RNA ligase	None

EP 3 907 296 A1

[0077] The ligation products provided by the PCR amplification as described in Table 22 are the ligation products by further preparing the extension products provided according to the systems in Table 3 and 4 according to the ligation reaction system in Tables 8 to 10. The specific parameters of the purification method used after ligation is disclosed in Table 11. The cfDNA samples referred to in the sequencing results in Table 23 are the QC samples. In the test of mixing mutant DNA into cfDNA at different ratio in Table 24, for the same condition number, except that DNA of NCI-H1975 cell line are mixed into cfDNA in different amount, the experimental steps are the same as in Table 22.

5

10

15

20

25

30

35

40

45

50

55

Table 23. Results of sequencing analysis

Condition Number	1	2	3	4	5	6	7	8	9	10	11
Total reads	367842	245740	284632	345320	299538	327782	243698	288673	296643	308712	312156
Reads aligning to hg19	184	211600	2343	327080	275605	1052	30292	126641	70809	270555	35336
Alignment rate	0.05%	86.11%	0.82%	94.72%	92.01%	0.32%	12.43%	43.87%	23.87%	87.64%	11.32%
Reads from the target region	61	125072	1039	192654	144252	402	14004	67436	34965	149671	16247
Ratio of reads from the target region	33.40%	59.11%	44.37%	58.90%	52.34%	38.20%	46.23%	53.25%	49.38%	55.32%	45.98%

Table 24. Results of detection of mutant DNA in cfDNA under different mixing rates

Condition Number	1	2	3	4	5	6	7	8	9	10	11
0 mixed	0	0	0	0	0	0	0	0	0	0	0
0.01% mixed	0	0.012%	0	0.021%	0.009%	0	0	0	0	0.019%	0
0.1% mixed	0	0.13%	0	0.12%	0.14%	0	0.08%	0.11%	0.09%	0.12%	0.08%
1% mixed	0	1.15%	0.48%	1.08%	0.83%	0	1.12%	1.13%	1.11%	1.18%	1.13%
10% mixed	2.30%	10.25%	9.53%	10.22%	9.89%	5.21 %	10.13%	10.23%	9.54%	10.22%	10.15%

[0078] It can be seen from Table 23 and Table 24, under the reaction conditions numbered 2, 4, 5, and 10, the sequencing alignment rates and the sensitivities are high, and the mutation can be correctly detected in the cfDNA sample with a mixing ratio of as low as 0.01%. The results showed that the blocking modification of the 3' terminal hydroxyl group of the probes is essential to prevent self-ligation or ligating to the linker, which significantly reduces the background noise. After binding to the template, the substituent group of the probe can be excised and the bound probe can effectively extend. The dNTP coupled with labeling molecule and the purification system thereof can significantly improve the purity of the target product in the final library, which is a plus to the technical effect of the present invention. In addition, the application of thermostable RNA ligase, the blocking of the 3' terminal of the linker DNA, and the purification after ligation can further improve the purity of the final library product and the sensitivity of detection, respectively/together constitute the technical solution of the present invention.

Embodiment 2

[0079] The oligonucleotide sequences used in embodiment 2 are shown in table 25.

Table 25

Name	SEQ ID NO:	Sequence (5'→3')
Probe 5	17	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TAGTTGGATGGGATTATTT-C3 Spacer
Probe 6	18	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TTTTTTTTTGGGAGTTTAAATAAAGATTA -C3 Spacer
Probe 7	19	<i>CCGCTTTCCTCTCTATGGGCAGTCGGTGAT</i> TTTAAAATAGAGTTAGTTTTAGTTTTT-C3 Spacer
ABar-X2	10	<i>PO₄-GGATCCNNNNNNGTTCTCCTTA</i> <i>CTGAGTCGGAGACACGCAGGGATGAGATGG</i> -C3 Spacer
AF	15	GTCTCAGCCTCTCTATGGGCAGTCGGTGAT
AR	16	CCATCTCATCCCTGCGTGTCTCCGACTCAG

[0080] The probes in this embodiment comprise a universal sequence, that is, bases shown in *italic*, which is the P1 sequence of the Ion Torrent system. The DNA sample used in this embodiment has been treated with bisulfite. Probe 5 (SEQ ID NO 17) targets a methylation region of the gene *SEPT9*, Probe 6 (SEQ ID NO: 18) targets a methylation region of the gene *NDRG4*, and Probe 7 (SEQ ID NO: 19) targets a methylation region of the gene *BMP3*.

[0081] The linker ABar-X2 (SEQ ID NO. 10) comprises a universal sequence, that is, bases shown in *italic*, which is the A sequence of the Ion Torrent system. The sequence underlined is a barcode, which can be replaced with different barcodes in different samples in the present invention. The "N" sequence is a molecular tag sequence (UMI). Among them, NNNNNN is a random sequence that is used to label different extension product molecules in the same sample. The 5' terminal of the linkers ABar-X2 (SEQ ID NO. 10) is phosphorylated.

[0082] AF (SEQ ID NO.15) and AR (SEQ ID NO.16) are the PCR pre-primers and post-primers (front primer and rear primer) used to amplify the library, respectively. The sequences thereof were matching the universal sequences (that is, the PI sequence and the A sequence) in the library molecule.

Reagents and Materials

[0083] The bisulfite conversion kit was purchased from Promega. The Agencourt AMPure magnetic beads were purchased from Beckman. The other main reagents used were the same as in embodiment 1.

Step 1. Preparation of the sample

[0084] Five samples of healthy individuals' plasma and intestinal cancer tissue were extracted, and the DNA samples were quantified by Qubit Fluorometer. The DNA from intestinal cancer tissues was sheared into fragments of about 160

EP 3 907 296 A1

bp and mixed into the plasma sample of healthy individuals at ratios of 10% and 5% to provide samples of 5% mixing rate and 10% mixing rate. Plasma sample from healthy individuals was used as a blank. Convert the plasma samples of healthy individuals using the Promega bisulfite conversion system. The converted DNA was quantified and formulated into a concentration of 1000 copies/ μL .

Step 2. Captured-extension

[0085] Prepare the captured-extension system according to table 26.

Table 26

Compositions	Volume (μL)	Final Concentration
2X High-fidelity DNA polymerase PCR Master Mix (with 3'-5' exonuclease activity)	25	1×
Probe 5 (10 μM)	0.5	100 nM
Probe 6 (10 μM)	0.5	100 nM
Probe 7 (10 μM)	0.5	100 nM
cfDNA samples	20	20000 copies
H ₂ O	3.5	/
Total	50	/

[0086] PCR reaction conditions of the captured-extension system is shown in table 27.

Table 27

Cycles	Temperature	Time
1 cycle	98°C	60 s
40 cycles	98°C	10 s
	62°C	20 s
	72°C	20 s
1 cycle	72°C	5 min

[0087] After completion of the captured-extension, the extension product was purified by silica gel column and eluted with 60 μL of the elution buffer. The captured-extension product in table 4 was further purified with the magnetic beads coated with streptavidin after the silica gel column purification, and finally dissolved in 60 μL of elution buffer.

Step 3. Single-stranded ligation

[0088] Perform the ligation reaction according to the ligation reaction system and conditions prepared in table 28.

Table 28

Compositions	Volume (μL)	Final Concentration	Ligation conditions
10X Ligase Buffer	4	1×	60°C 1 hour
50 mM MnCl ₂	2	2.5 mM	
5 M Betaine	8	1M	
Thermostable RNA ligase (100 U/(μL))	2	5 U/ μL	
ABar-X2(200 nM)	2	10 nM	
Extension product	8		
50% PEG	14	17.5%	

EP 3 907 296 A1

(continued)

Compositions	Volume (μL)	Final Concentration	Ligation conditions
Total	40	/	

[0089] After ligation, the ligation product was purified by silica gel column and finally eluted with 50 μL of elution buffer.

Step 4. Amplification of the ligation product(Library amplification)

[0090] According to the reaction system and conditions shown in table 29, PCR amplification of the ligation product was performed.

Table 29

Compositions		Volume
Ligation product		30 μL
5× Q5 Buffer		10 μL
dNTP (10mM each)		1 μL
Q5 high-fidelity DNA polymerase (2 U/ (μL)		1 μL
AF (10 μM)		1.5 μL
AR (10 μM)		1.5 μL
ddH ₂ O		5 μL
Total Volume		50 μL
Temperature(°C)	Time(s)	Cycles
98	30	1
98	10	12
72	30	
72	300	1

[0091] Purify the amplified product with 80 μL of Agencourt AMPure magnetic beads, and dissolve the purified product in 30 μL of elution buffer. At this point, the library construction for sequencing is completed.

Step 5. Sequencing and data analysis

[0092] Sequencing the prepared library using an Ion Proton system. The related operations includes water-in-oil PCR, library enrichment, biochip loading, and sequencing. The detailed procedure is described in the instruction manuals of Ion PITM Hi-QTM OT2 200 kit and Ion PITM Hi-QTM Sequencing 200 kit.

[0093] The analysis of sequencing results is shown in table 30 and table 31, respectively.

Table 30. Sequencing results 1

Type of sample	Total reads (M)	Reads from the target region (M)	Total alignment rate	Total specificity of the reads from the target region
Blank (QC)	1.261	0.448	74.35%	47.84%
5% mixed	1.376	0.491	79.78%	44.74%
10% mixed	1.508	0.513	75.42%	45.10%

Table 31.

Methylation degree of each gene under different mixing ratios of DNA from intestinal cancer tissues				
Name of genes	Methylation sites on genes	Blank	5% mixed	10% mixed
SEPT9	chr17: 75369623	1.89%	4.03%	5.82%
	chr17: 75369630	1.52%	5.04%	6.18%
	chr17: 75369657	1.35%	4.97%	5.55%
	chr17: 75369663	0.65%	3.67%	5.50%
BMP3	chr4: 81952061	3.01%	3.10%	6.63%
	chr4: 81952065	3.72%	4.22%	6.86%
	chr4: 81952078	2.48%	2.69%	5.99%
	chr4: 81952099	2.72%	3.57%	5.67%
NDRG4	chr16: 58497230	1.66%	2.92%	4.74%
	chr16: 58497236	1.41%	3.69%	4.19%
	chr16: 58497239	1.43%	3.00%	4.67%
	chr16: 58497251	3.17%	5.81%	5.89%
	chr16: 58497365	3.15%	5.88%	6.48%

[0094] It can be seen from table 30 and table 31, the samples with 5% and 10% intestinal cancer tissue DNA mixed were significantly different from the blank samples in the degree of methylation. The degree of methylation of the samples with 10% intestinal cancer tissue DNA mixed is highest, the blank sample had the lowest methylation degree, which was same with expectations. The results show that, the target region of DNA fragmented samples treated with bisulfite can also be well enriched and can be applied to the analysis of degree of gene methylation and related diseases by using the enrichment method of the present invention.

Embodiment 3

[0095] The oligonucleotide sequences used in embodiment 3 and 4 are shown in table 32:

Table 32

Name	SEQ NO.	Sequence (5'→3')
F-SRY	20	TTCCAGGAGGCACAGAAATTAC
R-SRY	21	CTTCCGACGAGGTCGATACT
MGB-SRY	22	CATGCACAGAGAGAAATACCCGAAT
F858	23	CAGATTTTGGGCG
ER1	13	TTTGCCTCCTTCTGC
EM1	14	AACTGCTGGGTGCGGA
F787	24	GCGTGATGAGT
R787	25	GCGTGGACAACCCCCAC
MGB787	26	CACGGTGGAGGTGAGGC
Uni-R	27	CCATCTCATCCCTGCGT
Uni-MGB	28	TCCGACTCAGTAAGGAGAACGA
Probe 8	29	CCGCTTTCTCTCTATGGGCAGTCGGTGATCCAATATTGTCTTT GTGTTCCCGGACATAGTCCT-C3 Spacer

(continued)

Name	SEQ NO.	Sequence (5'→3')
Probe 9	30	CCGCTTTCCTCTCTATGGGCAGTCGGTGATCTATGGCCATTCTT CCAGGAGGCACAGAAATTACT-C3 Spacer
ABar-X2	10	PO4- GGATCCNNNNNNGTTCTCCTTACTGAGTCGGAGACACGCAGGG ATGAGATGG -C3 Spacer

[0096] Main reagents and materials:

Same as in Embodiment 1, the calibrator used in qPCR was prepared according to the conventional method of molecular cloning.

Step 1. Sample preparation

[0097] The non-small cell lung cancer cell line NCI-H1975 contains both the mutation of EGFR L858R site and SNP of Q787 site. Genomic DNA of non-small cell lung cancer cell line NCI-H1975 and healthy human leukocytes was extracted using XX kit. The genomic DNA is broken to be with molecular weight range of 100-300bp by ultrasound. After quantification by the Qubit fluorescence quantifier, the genomic DNA of NCI-H1975 was mixed into the genomic DNA of healthy human leukocytes at a ratio of 1%, 0.1%, 0.03%, 0.01%, 0%, and captured-extension, ligation and library expansion were applied to the samples by step 2 to step 6 in embodiment 1 of the present invention.

Step 2. qPCR detection of different specific primer/probe' combination

[0098] The trispecific, bispecific and monospecific detection systems for the EGFR gene L858R site are shown in tables 33-35.

Table 33. Trispecific detection system for L858R site

Reagents	Dosage (μL)	Final concentration
2×Taqman Mix	10	1×
F858(10μM)	0.6	300nM
R858(10μM)	0.6	300nM
MGB858(10μM)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

Table 34. Bispecific detection system for L858R site

Reagents	Dosage (μL)	Final concentration
2×Taqman Mix	10	1×
F858(10μM)	0.6	300nM
Uni-R(10μM)	0.6	300nM
MGB858(10μM)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/

EP 3 907 296 A1

(continued)

Reagents	Dosage (μ L)	Final concentration
Total	20	/

Table 35. Monospecific detection system for L858R site

Reagents	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F858(10 μ M)	0.6	300nM
Uni-R(10 μ M)	0.6	300nM
Uni-MGB(10 μ M)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

[0099] The trispecific, bispecific, and monospecific detection systems for EGFR gene Q787Q site are shown in tables 36-38.

Table 36. Trispecific detection system for Q787Q site:

Reagents	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F787(10 μ M)	0.6	300nM
R787(10 μ M)	0.6	300nM
MGB787(10 μ M)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

Table 37. Bispecific detection system for Q787Q site:

Reagents	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F787(10 μ M)	0.6	300nM
Uni-R (10 μ M)	0.6	300nM
MGB787(10 μ M)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

Table 38. Monospecific detection system for Q787Q site:

Reagents	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$

(continued)

Reagents	Dosage (μL)	Final concentration
F787(10 μM)	0.6	300nM
Uni-R(10 μM)	0.6	300nM
Uni-MGB(10 μM)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

qPCR detection procedures are shown in table 39.

Table 39

Cycle	Temperature	Time
1 cycle	95°C	4 min
40 cycles	95°C	10 s
	61°C	30 s

Experimental results and analysis

[0100] Results 1. The detection results of L858R sites with combinations of different specific primer/probe are shown in tables 40–42, respectively. Among them, the trispecific, bispecific and monospecific detection systems correspond to the reaction systems in tables 33–35, respectively.

Table 40. Results of trispecific detection of L858R site ($\Delta\text{CT}=\text{ABS}(\text{mean Qi-Q5})$)

Mixing rate	CT1	CT2	CT Mean	ΔCT
1%	26.13	26.49	26.31	11.715
0.1%	30.35	30.56	30.46	7.570
0.03%	33.43	33.85	33.64	4.385
0.01%	37.34	36.52	36.93	1.095
0	38.52	37.53	38.03	0

Table 41. Results of bispecific detection of L858R site

Mixing rate	CT1	CT2	CT Mean	ΔCT
1%	24.12	24.14	24.13	13.040
0.1%	28.73	28.24	28.49	8.685
0.03%	30.82	31.32	31.07	6.100
0.01%	33.43	34.18	33.81	3.365
0	37.32	37.02	37.17	0

Table 42. Results of monospecific detection of L858R site

Mixing rate	CT1	CT2	CT Mean	ΔCT
1%	23.64	23.52	23.58	12.850

(continued)

Mixing rate	CT1	CT2	CT Mean	Δ CT
0.1%	27.38	27.02	27.20	9.230
0.03%	29.24	29.87	29.56	6.875
0.01%	32.29	31.58	31.94	4.495
0	36.24	36.62	36.43	0

[0101] Results 2. The detection results of Q787Q sites with combinations of different specific primer/probe are shown in tables 43-45 respectively. Among them, the trispecific, bispecific and monospecific detection systems correspond to the reaction systems in tables 36-38, respectively.

Table 43. Trispecific test results of Q787Q site

Mixing rate	CT1	CT2	CT Mean	Δ CT
1%	26.09	25.85	25.97	11.055
0.1%	30.52	30.82	30.67	6.355
0.03%	34.07	33.59	33.83	3.195
0.01%	36.82	36.43	36.63	0.400
0	37.14	36.91	37.03	0

Table 44. Bispecific detection results of Q787Q site

Mixing rate	CT1	CT2	CT Mean	Δ CT
1%	25.45	25.48	25.47	10.820
0.1%	29.01	29.48	29.25	7.040
0.03%	31.77	32.11	31.94	4.345
0.01%	34.12	34.98	34.55	1.735
0	36.05	36.52	36.29	0

Table 45. Monospecific detection results of Q787Q site

Mixing rate	CT1	CT2	CT Mean	Δ CT
1%	24.44	24.17	24.31	11.950
0.1%	27.84	28.19	28.02	8.240
0.03%	30.45	31.01	30.73	5.525
0.01%	32.98	32.45	32.72	3.540
0	36.03	36.48	36.26	0

[0102] Conclusion: For reference products with different mixing rates, bispecific and monospecific primer/probe systems have a larger Δ CT value and a higher effective resolution than trispecific system. when Δ CT=3 is used as cut off, for the L858R site, the detection limits of the monospecific system, bispecific system and trispecific system are 0.01%, 0.01% and 0.03% respectively; for Q787 site, the detection limits are 0.01%, 0.03% and 0.01%, respectively. This shows that the monospecific system and the bispecific system still have a reliable detection rate even the mixing rate is as low as 0.03%.

Embodiment 4.

[0103] The oligonucleotide sequences used in this embodiment are shown in table 32.

[0104] Main reagents and materials: same as embodiment 3.

Step 1. Sample preparation

[0105] Extract plasma free DNA of male and female respectively by plasma free nucleic acid extraction kit. After quantification with Qubit fluorescence quantifier, at the ratio of 1%, 0.1%, 0.03%, 0.01%, 0%, mix male cfDNA into the female cfDNA sample using the female cfDNA as a blank control. Perform steps 2~6 in embodiment 1 of the present invention to apply captured-extension, ligation and library amplification to the sample.

Step 2. qPCR detection of different specific primer/probe's combinations

[0106] Trispecific, bispecific, and monospecific detection systems for specific SRY gene of male are shown in tables 46-48.

Table 46. Trispecific detection system for male specific SRY gene:

Reagent	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY(10 μ M)	0.6	300nM
R-SRY(10 μ M)	0.6	300nM
MGB-SRY(10 μ M)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

Table 47. Bispecific detection system for male specific SRY gene:

Reagent	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY(10 μ M)	0.6	300nM
Uni-R(10 μ M)	0.6	300nM
MGB-SRY(10 μ M)	0.2	100nM
H ₂ O	6.6	/
Calibrator or sample to be tested	2	/
Total	20	/

Table 48. Monospecific detection system for male specific SRY gene:

Reagent	Dosage (μ L)	Final concentration
2 $\times$ Taqman Mix	10	1 $\times$
F-SRY(10 μ M)	0.6	300nM
Uni-R(10 μ M)	0.6	300nM
Uni-MGB (10 μ M)	0.2	100nM
H ₂ O	6.6	/

(continued)

Reagent	Dosage (μL)	Final concentration
Calibrator or sample to be tested	2	/
Total	20	/

qPCR detection procedure is shown in table 49.

Table 49.

Cycles	Temperature	Time
1 cycle	95°C	4 min
40 cycles	95°C	10 s
	61°C	30 s

Test results and analysis

[0107] Results: The test results of male specific SRY genes with different specific primer/probe's combinations are shown in tables 50-52, respectively. The trispecific, bispecific and monospecific detection systems correspond to the reaction systems in tables 46-48, respectively.

Table 50. Trispecific test results of male specific SRY gene ($\Delta CT = ABS(\text{mean } Q_i - Q_5)$)

Mixing rate	CT1	CT2	CT mean	ΔCT
1%	31.23	32.08	31.66	13.345
0.1%	35.71	35.94	35.83	9.175
0.03%	37.85	38.02	37.94	7.065
0.01%	39.25	39.52	39.39	5.615
0	ND	ND	45.00	0

Table 51. Bispecific test results of male specific SRY gene

Mixing rate	CT1	CT2	CT mean	ΔCT
1%	30.55	30.99	30.77	14.230
0.1%	34.23	34.98	34.61	10.395
0.03%	36.32	37.94	37.13	7.870
0.01%	39.48	39.12	39.30	5.700
0	ND	ND	45.00	0

Table 52. Monospecific test results of male specific SRY gene

Mixing rate	CT1	CT2	CT mean	ΔCT
1%	29.38	29.68	29.53	15.470
0.1%	33.28	33.89	33.59	11.415
0.03%	35.45	35.82	35.64	9.365
0.01%	38.21	37.89	38.05	6.950
0	ND	ND	45.00	0

[0108] Conclusion: As a gene only on Y chromosome, SRY has high detection sensitivity by the three methods. If $\Delta CT=3$ is used as the cut off, the monospecific, bispecific, and trispecific systems each can achieve a detection sensitivity of 0.01%. However, judging from the specific CT value, the detection rate of the monospecific system is higher.

[0109] In summary, the present invention effectively overcomes various shortcomings in the prior art and has high industrial utilization value.

[0110] The above embodiments are only illustrative of the principles and effects of the present invention, not intended to limit the present invention. Anyone familiar with this technology can modify or change the above embodiments without departing from the spirit and scope of the present invention. Therefore, all equivalent modifications or changes made by those with common knowledge in the technical field without departing from the spirit and technical idea disclosed by the present invention should still be covered by the claims of the present invention.

EP 3 907 296 A1

SEQUENCE LISTING

<110> SHANGHAI ZENISIGHT LTD.
 5 <120> GENE TARGET REGION ENRICHMENT METHOD AND KIT
 <130> N422244EP
 <140> EP 19907784.3
 10 <141> 2019-12-09
 <150> CN20191002408
 <151> 2019-01-02
 <150> CN201910897689
 15 <151> 2019-09-20
 <160> 30
 <170> PatentIn version 3.5
 20 <210> 1
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 25 <220>
 <223> Synthetic polynucleotide - Probe 1
 <400> 1
 ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa 57
 30 <210> 2
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 35 <220>
 <223> Synthetic polynucleotide - Probe 2
 <400> 2
 ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa 57
 40 <210> 3
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 45 <220>
 <223> Synthetic polynucleotide - Probe 2b
 <400> 3
 ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa 57
 50 <210> 4
 <211> 57
 <212> DNA
 55 <213> Artificial Sequence
 <220>

<223> Synthetic polynucleotide - Probe 2c

<400> 4

ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa

57

<210> 5

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 2n

<400> 5

ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa

57

<210> 6

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 2d

<400> 6

ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgaa

57

<210> 7

<211> 57

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 3

<400> 7

ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tggtgat

57

<210> 8

<211> 58

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 4

<400> 8

ccgctttcct ctctatgggc agtcggtgat cctggcagcc aggaacgtac tgrgtgat

58

<210> 9

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - ABar-X1

<220>
 <221> misc_feature
 <222> (7)..(12)
 <223> n is a, c, g, or t
 5
 <400> 9
 ggatccnnnnn nncagcttgg actgagtcgg agacacgcag ggatgagatg g 51

 <210> 10
 <211> 52
 <212> DNA
 <213> Artificial Sequence
 10

 <220>
 <223> Synthetic polynucleotide - ABar-X2
 15

 <220>
 <221> misc_feature
 <222> (7)..(12)
 <223> n is a, c, g, or t
 20

 <400> 10
 ggatccnnnnn nngttctcct tactgagtcg gagacacgca gggatgagat gg 52

 <210> 11
 <211> 51
 <212> DNA
 <213> Artificial Sequence
 25

 <220>
 <223> Synthetic polynucleotide - ABar-X3
 30

 <220>
 <221> misc_feature
 <222> (7)..(12)
 <223> n is a, c, g, or t
 35

 <400> 11
 ggatccnnnnn nnttacctta gctgagtcgg agacacgcag ggatgagatg g 51
 40

 <210> 12
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 45

 <220>
 <223> Synthetic polynucleotide - EF-1

 <400> 12
 gatcacagat tttgggc 17
 50

 <210> 13
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 55

 <220>

	<223> Synthetic polynucleotide - ER-1	
5	<400> 13 tttgcctcct tctgc	15
	<210> 14 <211> 16 <212> DNA <213> Artificial Sequence	
10	<220> <223> Synthetic polynucleotide - EM-1	
15	<400> 14 aactgctggg tgcgga	16
	<210> 15 <211> 30 <212> DNA <213> Artificial Sequence	
20	<220> <223> Synthetic polynucleotide - AF	
25	<400> 15 gtctcagcct ctctatgggc agtcggtgat	30
	<210> 16 <211> 30 <212> DNA <213> Artificial Sequence	
30	<220> <223> Synthetic polynucleotide - AR	
35	<400> 16 ccatctcatc cctgcgtgtc tccgactcag	30
	<210> 17 <211> 49 <212> DNA <213> Artificial Sequence	
40	<220> <223> Synthetic polynucleotide - Probe 5	
45	<400> 17 ccgctttcct ctctatgggc agtcggtgat tagttggatg ggattattt	49
	<210> 18 <211> 58 <212> DNA <213> Artificial Sequence	
50	<220> <223> Synthetic polynucleotide - Probe 6	
55	<400> 18	

	ccgctttcct ctctatgggc agtcggtgat ttttttttgg gagtttaaataaagatta	58
5	<210> 19 <211> 57 <212> DNA <213> Artificial Sequence	
10	<220> <223> Synthetic polynucleotide - Probe 7	
	<400> 19 ccgctttcct ctctatgggc agtcggtgat tttaaaatag agttagtttt agttttt	57
15	<210> 20 <211> 22 <212> DNA <213> Artificial Sequence	
20	<220> <223> Synthetic polynucleotide - F-SRY	
	<400> 20 ttccaggagg cacagaaatt ac	22
25	<210> 21 <211> 20 <212> DNA <213> Artificial Sequence	
30	<220> <223> Synthetic polynucleotide - R-SRY	
35	<400> 21 cttccgacga ggtagatact	20
40	<210> 22 <211> 25 <212> DNA <213> Artificial Sequence	
	<220> <223> Synthetic polynucleotide - MGB-SRY	
45	<400> 22 catgcacaga gagaaataacc cgaat	25
50	<210> 23 <211> 13 <212> DNA <213> Artificial Sequence	
55	<220> <223> Synthetic polynucleotide - F858	
	<400> 23 cagatttttg gcg	13

<210> 24
 <211> 11
 <212> DNA
 <213> Artificial Sequence
 5
 <220>
 <223> Synthetic polynucleotide - F787
 <400> 24
 10 gcgtgatgag t 11
 <210> 25
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 15
 <220>
 <223> Synthetic polynucleotide - R787
 <400> 25
 20 gcgtggacaa cccccac 17
 <210> 26
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 25
 <220>
 <223> Synthetic polynucleotide - MGB787
 <400> 26
 30 cacggtggag gtgaggc 17
 <210> 27
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 35
 <220>
 <223> Synthetic polynucleotide - Uni-R
 <400> 27
 40 ccatctcatc cctgcgt 17
 <210> 28
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 45
 <220>
 <223> Synthetic polynucleotide - Uni-MGB
 <400> 28
 50 tccgactcag taaggagaac ga 22
 <210> 29
 <211> 64
 <212> DNA
 55

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 8

<400> 29

ccgctttcct ctctatgggc agtcggtgat ccaatattgt ctttgtgttc ccggacatag 60

tcct 64

<210> 30

<211> 65

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide - Probe 9

<400> 30

ccgctttcct ctctatgggc agtcggtgat ctatggccat tcttccagga ggcacagaaa 60

ttact 65

Claims

1. A method for enriching a target region of a gene, comprising:

- (1) amplifying fragmented DNA comprising the target region via specific probes to provide captured-extension products, wherein the specific probe comprises a sequence complementary to the target region of the fragmented DNA, and the 3' terminal nucleotide of the specific probe is modified to prevent a ligation reaction at the 3' terminal of the specific probe;
- (2) linking the 3' terminal of the captured-extension product obtained in step (1) to linker DNA to obtain a ligation product.

2. The method for enriching a target region of a gene according to claim 1, wherein in step (1), the fragmented DNA comprises double-stranded DNA, single-stranded DNA, and cDNA, and length of the fragmented DNA ranges from 25~200 bp; and/or, the amplification system of the step (1) includes specific probes, DNA polymerase and dNTP.

3. The method for enriching a target region of a gene according to claim 2, wherein

- the DNA polymerase has a 3' - 5' exonuclease activity; and/or,
the dNTP is further coupled with a labeling molecule, preferably, the labeling molecule is biotin.

4. The method for enriching a target region of a gene according to claim 1, wherein

- the amplification system of step (1) further comprises an active substance, configured to cleave the substituent group at the 3' terminal of the specific probe after binding to the target region; preferably, the active substance is a nuclease; and/or,
the specific probe further comprises a universal sequence able to be recognized by a sequencing system; and/or,
the 3-position hydroxyl group of the 3' terminal nucleotide of the specific probe is substituted; and/or,
the substituent group of the 3' terminal of the specific probe is selected from the group consisting of a hydrogen atom, a C3 Spacer, a C6 Spacer, a phosphate group, and an amino group; and/or,
the tail region at the 3' terminal of the specific probe contains mismatched bases.

5. The method for enriching a target region of a gene according to claim 1, wherein

a single-stranded ligase is further involved in step (2), and said single-stranded ligase is a T4 RNA ligase or a thermostable RNA ligase; and/or,
 5 the 5' terminal nucleotide of the linker DNA is modified, and the linker DNA has a single-stranded structure at the reaction temperature in step (2); preferably, the substituent group of the 5' terminal nucleotide of the linker DNA is substituted by a phosphate group or an adenosine group; and/or,
 the linker DNA has a partially double-stranded structure with a sticky end at the 5' end region; and/or,
 10 the linker DNA comprises one or more of the following: a universal sequence able to be recognized by a sequencing system, and a sample tag sequence, and a molecular tag sequence.

6. The method for enriching a target region of a gene according to claim 1, wherein

the step (1) further comprises purification of the captured-extension product by magnetic bead; and/or,
 15 the step (2) further comprises purification of the ligation product by silica gel column and/or heat treatment.

7. The method for enriching a target region of a gene according to claim 1, wherein further comprising
 (3) amplifying the ligation product provided from step (2), preferably, in step (3), the PCR amplification primer
 20 comprises sequences matching the universal sequences of the specific probe and/or universal sequences of the linker DNA.

8. The method according to claim 7, wherein further comprising:
 (4) sequencing of the amplified ligation product to provide sequencing results of the target regions.

- 25 9. The method for enriching a target region of a gene according to claim 1, further comprising:

(5) detecting the ligation product provided from step (2) with a detection primer 1, a detection primer 2, and a probe 3,
 wherein
 30 at least one of the above detection primer 1, detection primer 2, and probe 3 contains a gene-specific sequence; preferably, the detection primer 1 comprises a gene-specific sequence, and the detection primer 2 and probe 3 comprise universal sequences; and/or,
 the detection primer 2 and/or probe 3 comprise gene-specific sequences, either;
 preferably, the probe 3 comprises a labeling molecule, and the sequence of the probe 3 is not complementary
 35 to that of detection primer 1 or detection primer 2.

10. The method for enriching a target region of a gene according to claim 1, wherein the method is used for gene detection.

11. A kit for enriching target region of fragmented DNA, comprising specific probes and linker DNA suitable for method
 40 for enriching a target region of a gene according to any one of claims 1-9.

12. The kit according to claim 11, further comprising one or more of the following:

RNA ligase, dNTP coupled with labeling molecule, DNA polymerase, and nuclease; and/or,
 45 further comprises a forward primer and a reverse primer, the forward primer and the reverse primer comprising a sequence that is at least partially complementary to the universal sequence of the specific probe and the universal sequence of the linker DNA.

13. The kit according to claim 12, further comprising detection primer 1, detection primer 2 and probe 3, at least one of
 50 which contains gene-specific sequences.

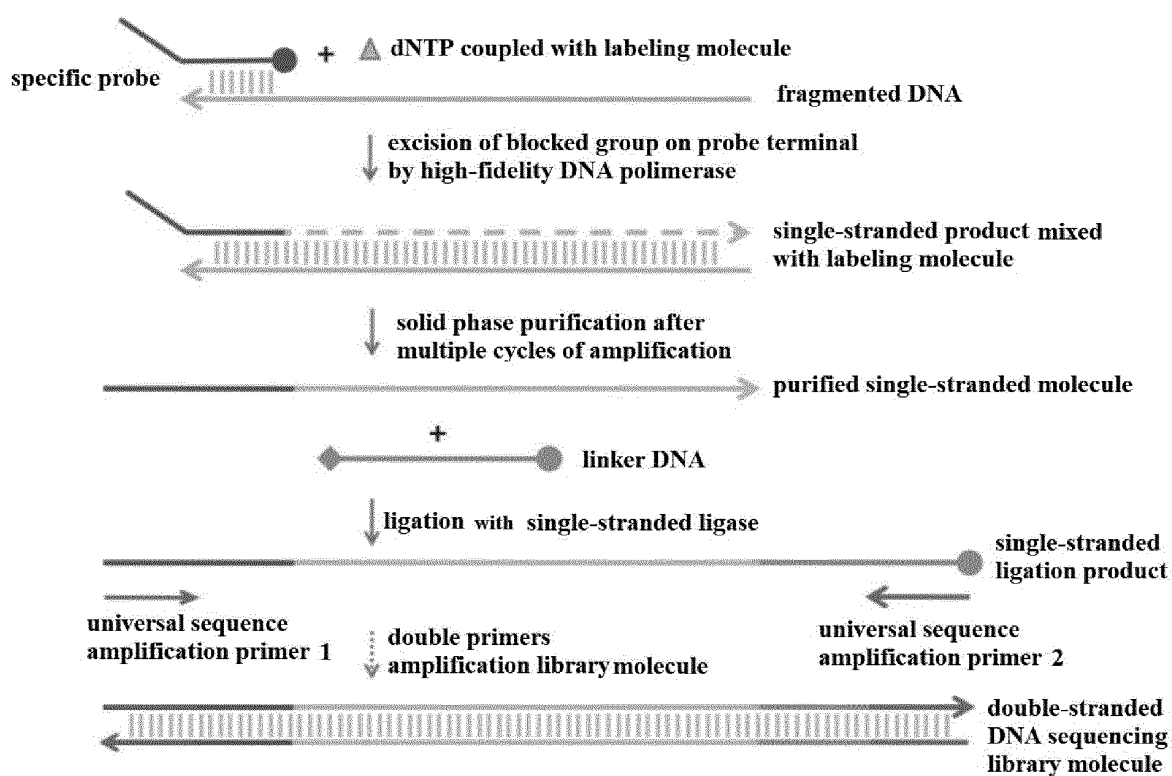


FIG.1

Universal sequence for probe sequencing (P1)	Complementary sequence of probe	Sequence of target region	Sequence of linker	Molecular tag	Sample tag	Universal sequence for linker sequencing (A)
---	---------------------------------------	---------------------------------	-----------------------	------------------	---------------	--

FIG.2

AF sequence	Exon sequence of EGFR gene 21	AR sequence
-------------	-------------------------------	-------------

FIG.3

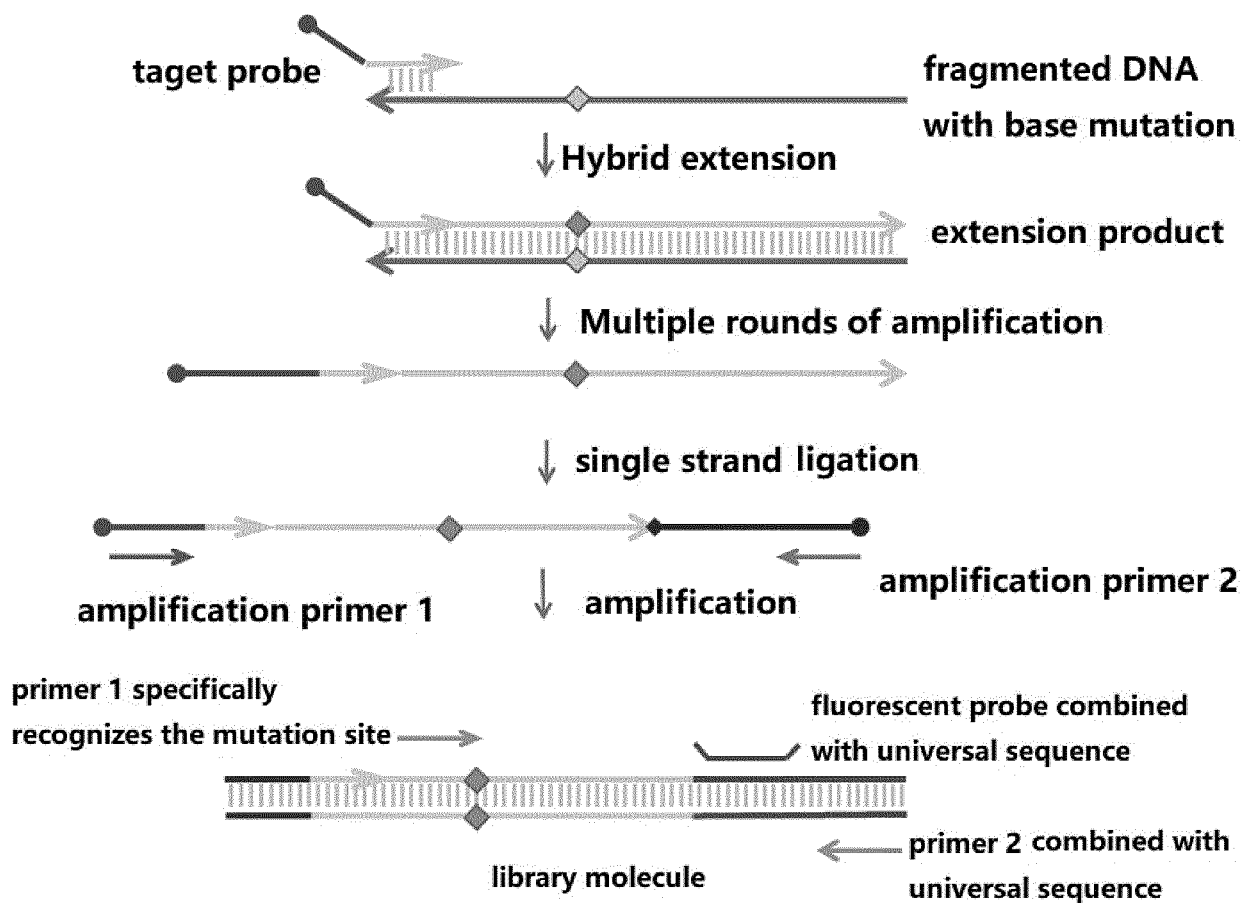


FIG.4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/123891

A. CLASSIFICATION OF SUBJECT MATTER C12Q 1/6806(2018.01)i According to International Patent Classification (IPC) or to both national classification and IPC												
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) C12Q Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched												
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) CNMED, CNABS, CPEA, DWPI, SIPOABS, AUABS, TWMED, ILABS, TWABS, HKABS, MOABS, SGABS, CNKI, NCBI, EMBL, GoogleScholar: 富集, 捕获延伸, 探针, 扩增 capture-extension product CEP, Specific probe hybridization capture, Probe capture based enrichment												
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>CN 107922970 A (F.HOFFMANN-LA ROCHE AG) 17 April 2018 (2018-04-17) see entire document, especially the abstract</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECHNOLOGY CO., LTD.) 15 February 2018 (2018-02-15) see entire document, especially the abstract</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>Johannes Dapprich et al. "The Next Generation of Target Capture Technologies-large DNA Fragment Enrichment and Sequencing Determines Regional Genomic Variation of High Complexity" <i>BMC Genomics</i>, Vol. 17, 09 July 2016 (2016-07-09), the abstract, and figure 1</td> <td>1-13</td> </tr> </tbody> </table>	Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	A	CN 107922970 A (F.HOFFMANN-LA ROCHE AG) 17 April 2018 (2018-04-17) see entire document, especially the abstract	1-13	A	WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECHNOLOGY CO., LTD.) 15 February 2018 (2018-02-15) see entire document, especially the abstract	1-13	A	Johannes Dapprich et al. "The Next Generation of Target Capture Technologies-large DNA Fragment Enrichment and Sequencing Determines Regional Genomic Variation of High Complexity" <i>BMC Genomics</i> , Vol. 17, 09 July 2016 (2016-07-09), the abstract, and figure 1	1-13
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.										
A	CN 107922970 A (F.HOFFMANN-LA ROCHE AG) 17 April 2018 (2018-04-17) see entire document, especially the abstract	1-13										
A	WO 2018028001 A1 (CHENGDU NUOEN BIOLOG TECHNOLOGY CO., LTD.) 15 February 2018 (2018-02-15) see entire document, especially the abstract	1-13										
A	Johannes Dapprich et al. "The Next Generation of Target Capture Technologies-large DNA Fragment Enrichment and Sequencing Determines Regional Genomic Variation of High Complexity" <i>BMC Genomics</i> , Vol. 17, 09 July 2016 (2016-07-09), the abstract, and figure 1	1-13										
☐ Further documents are listed in the continuation of Box C. ☒ See patent family annex. * Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family												
Date of the actual completion of the international search 12 February 2020	Date of mailing of the international search report 06 March 2020											
Name and mailing address of the ISA/CN China National Intellectual Property Administration (ISA/CN) No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China Facsimile No. (86-10)62019451	Authorized officer Telephone No.											

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/123891

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2019/123891

5

10

15

20

25

30

35

40

45

50

55

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	107922970	A	17 April 2018	WO	2017021449	A1	09 February 2017
				EP	3332024	A1	13 June 2018
				US	2017037459	A1	09 February 2017
				JP	2018521675	A	09 August 2018
				CA	2994601	A1	09 February 2017
WO	2018028001	A1	15 February 2018	CN	106282161	A	04 January 2017

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Non-patent literature cited in the description

- **SAMBROOK et al.** MOLECULAR CLONING : A LABORATORY MANUAL. Cold Spring Harbor Laboratory Press, 1989 [0034]
- MOLECULAR CLONING : A LABORATORY MANUAL. 2001 [0034]
- **AUSUBEL et al.** CURRENT PROTOCOLS IN MOLECULAR BIOLOGY. John Wiley & Sons, 1987 [0034]
- METHODS IN ENZYMOLOGY. Academic Press [0034]
- **WOLFFE.** CHROMATIN STRUCTURE AND FUNCTION. Academic Press, 1998 [0034]
- METHODS IN ENZYMOLOGY. Chromatin. Academic Press, 1999, vol. 304 [0034]
- Chromatin Protocols. METHODS IN MOLECULAR BIOLOGY. Humana Press, 1999, vol. 119 [0034]